

---

# Hilbert Space Embeddings of PSRs

---

**Byron Boots**

Machine Learning Dept.  
Carnegie Mellon University  
Pittsburgh, PA  
beb@cs.cmu.edu

**Arthur Gretton**

Gatsby Unit  
University College London  
London, UK  
arthur.gretton@gmail.com

**Geoffrey J. Gordon**

Machine Learning Dept.  
Carnegie Mellon University  
Pittsburgh, PA  
ggordon@cs.cmu.edu

## 1 Introduction and Related Work

Many problems in machine learning and artificial intelligence involve discrete-time partially observable nonlinear dynamical systems. If the observations are discrete, then *Hidden Markov Models* (HMMs) (Rabiner, 1989) or, in the control setting, *Partially Observable Markov Decision Processes* (POMDPs) (Sondik, 1971) can be used to represent belief as a discrete distribution over latent states. *Predictive State Representations* (PSRs) (Littman et al., 2002) are generalizations of POMDPs that have attracted interest because they often have greater representational capacity for a fixed model dimension. In contrast to the latent-variable representations of POMDPs, PSRs represent the state of a dynamical system by tracking occurrence probabilities of a set of future events (called *tests*) conditioned on past events (called *histories*). Because tests and histories are observable quantities, it has been suggested that learning PSRs should be easier than learning POMDPs.

Recently, (Boots et al., 2010) proposed a *spectral* algorithm for learning PSRs with discrete observations and actions. At its core, the algorithm performs a singular value decomposition of a matrix of joint probabilities of tests and partitions of histories, and then uses linear algebra to recover parameters which allow predicting, simulating, and filtering in the modeled system. Because PSRs of dynamical systems consist of expectations of *observable* quantities that can be estimated directly from training data, a statistically consistent estimate of the PSR can be found without resorting to local search. This is an important benefit of spectral learning algorithms for PSRs over heuristics (like EM) applied to latent variable representations.

Despite their positive properties, PSR models of controlled dynamical systems are usually restricted to discrete observations and actions with only moderate cardinality. For continuous actions and observations, and for actions and observations with large cardinalities, standard learning algorithms for PSRs often run into trouble: we cannot hope to see each or action observation more than a small number of times and we cannot gather enough data to estimate the PSR parameters accurately unless we make additional assumptions.

In this paper, we fully generalize PSRs to continuous observations and actions using a recent concept called Hilbert space embeddings of distributions (Smola et al., 2007; Sriperumbudur et al., 2008). The essence of our method is to represent distributions of tests, histories, observations, and actions, as points in (possibly) infinite-dimensional Reproducing Kernel Hilbert Spaces (RKHSs). During filtering we update these distributions using a kernel version of Bayes' rule (Fukumizu et al., 2011). To improve computational tractability, we develop a spectral *system identification* method to learn a succinct parameterization of the target system.

Our approach is similar to recent work that applies kernel methods to dynamical system modeling and reinforcement learning, which we summarize here. Song et al. (Song et al., 2010) proposed a nonparametric approach to learning HMM representations in RKHS. The resulting dynamical system model, called Hilbert Space Embeddings of Hidden Markov Models (HSE-HMMs), proved to be an impressive alternative to competing dynamical system models on several experimental benchmarks. Despite these successes, HSE-HMMs have two major limitations: first, the update rule for the HMM relies on *density estimation* instead of direct Bayesian inference in Hilbert space, which results in a theoretically awkward model. Second, the model lacks the capacity to reason about actions, which limits the scope of the algorithm. **Our model is a PSR extension of HSE-HMMs that adds control inputs and updates state using a kernelized version of Bayes' rule.**

Grünewälder et al. (Grunewalder et al., 2012) proposed a nonparametric approach to learning transition dynamics in Markov decision processes (MDPs) by representing the MDP transitions as conditional distributions in RKHS. This work was extended to POMDPs by Nishiyama et al. (Nishiyama et al., 2012). The resulting Hilbert space embedding of POMDPs represents distributions over the states, observations, and actions as embeddings in RKHS and uses kernel Bayes rule to update these distribution embeddings. Critically, the authors only provided results for fully observable models, where the training data includes labels for the true latent states. **By contrast, our algorithm only requires access to an (unlabeled) sequence of actions and observations and learns the more expressive PSR model.**

## 2 Predictive State Representations

The focus of this work is learning a special predictive representation of a dynamical system called a Predictive State Representation (PSR). PSRs represent state as a set of predictions of observable experiments or *tests* that one could perform in the system. Specifically, a test of length  $N_F$  is an ordered sequence of future action-observations pairs  $\tau_i = a_1, o_1, \dots, a_{N_F}, o_{N_F}$  that can be executed and observed at any time  $t$ . Likewise, a *history* is an ordered sequence of actions and observations  $h_t = a_1, o_1, \dots, a_{t-1}, o_{t-1}$  that has been executed and observed prior to a given time  $t$ . A test  $\tau_i$  is said to be *executed* at time  $t$  if we take the sequence of actions specified by the test  $\tau_i^A = a_1, \dots, a_{N_F}$ . It is said to *succeed* at time  $t$  if it is taken and if the sequence of observations in the test  $\tau_i^O = o_1, \dots, o_{N_F}$  matches the observations generated by the system. The *prediction* for test  $i$  at time  $t$  is the probability of the test succeeding given a history  $h_t$  and given that we take it. In this paper we assume that our data was generated by a *blind* policy, where future actions do not rely on future observations. This means that the PSR state is the *conditional* probability of observation sequences given action sequences:  $\mathbb{P}[\tau_{i,t}^O \mid \tau_{i,t}^A, h_t]$ . We write  $\tau(h_t)$  for the *prediction vector* of success probabilities for tests  $\tau_i$  given a history  $h_t$ , with elements:  $\tau_i(h_t) = \mathbb{P}[\tau_{i,t}^O \mid \text{do}(\tau_{i,t}^A), h_t]$ . The key idea behind a PSR is that if we know the expected outcomes of executing all possible tests, then we also know everything there is to know about the state of a dynamical system (Singh et al., 2004). Given that we take an action  $a_t$  and see a new observation  $o_t$ , the predictive density of tests are updated using Bayes' rule.

In this paper we will represent the predictive probability density for tests as an expectation of features in Hilbert space and use kernel Bayes' rule to update this density given new information. We introduce the necessary background in Section 3 below.

## 3 Hilbert Space Embeddings

Let  $\mathcal{F}$  be a reproducing kernel Hilbert space (RKHS) associated with kernel  $K_X(x, x') \stackrel{\text{def}}{=} \langle \phi^X(x), \phi^X(x') \rangle_{\mathcal{F}}$ . Following (Smola et al., 2007), we define the mean map of a probability distribution  $\mathbb{P}$  into RKHS  $\mathcal{F}$  to be  $\mu_X \stackrel{\text{def}}{=} \mathbb{E}[\phi^X(X)]$ , also called the Hilbert space embedding of  $\mathbb{P}$ . For all  $f \in \mathcal{F}$ ,  $\mathbb{E}[f(X)] = \langle f, \mu_X \rangle_{\mathcal{F}}$  by the reproducing property and linearity of expectations.

Given a joint distribution  $\mathbb{P}[X, Y]$  over two variables  $X$  on  $\mathcal{X}$  and  $Y$  on  $\mathcal{Y}$ <sup>1</sup> the *uncentered* covariance operator  $\mathcal{C}_{XY}$  is (Baker, 1973)  $\mathcal{C}_{XY} \stackrel{\text{def}}{=} \mathbb{E}_{XY}[\phi^X(X) \otimes \phi^Y(Y)]$ . Based on covariance operators (Song et al., 2009) define a conditional embedding operator which allows us to compute conditional expectations  $\mathbb{E}[\phi^Y(Y) \mid X]$  as linear operators in RKHS. We define the conditional embedding operator  $\mathcal{W}_{Y|X} \stackrel{\text{def}}{=} \mathcal{C}_{YX} \mathcal{C}_{XX}^{-1}$  such that for all  $g \in \mathcal{G}$ ,  $\mathbb{E}[g(Y) \mid x] = \langle g, \mathcal{W}_{Y|X} \phi^X(x) \rangle_{\mathcal{G}}$  given several assumptions.

Finally, we define Kernel Bayes' Rule (KBR) which performs a Bayes rule update  $\mathbb{P}[X \mid Y = y, Z = z] = \frac{\mathbb{P}[X, Y=y|Z=z]}{\mathbb{P}[Y=y|Z=z]}$  entirely in RKHS. We accomplish this by implementing Bayes' rule in terms of conditional covariance operators (Fukumizu et al., 2011):  $\mu_{X|y,z} = \mathcal{C}_{XY|z} \mathcal{C}_{YY|z}^{-1} \phi(y)$ .

In practice, the above operators can be expressed and estimated in terms of Gram matrices. Unfortunately, due to space constraints we do not go into the details here.

<sup>1</sup>We assume a kernel  $K_y(y, y') = \langle \phi^Y(y), \phi^Y(y') \rangle_{\mathcal{G}}$  is defined on  $\mathcal{Y}$  with associated RKHS  $\mathcal{G}$ .

## 4 Predictive Representations in RKHS

We will focus on learning the conditional embedding operator  $\mathcal{W}_{\mathcal{T}^\circ|\mathcal{T}^\mathcal{A},h_t}$  for the predictive density of tests  $\mathbb{P}[\tau_t^\circ | \tau_t^\mathcal{A}, h_t]$  of a PSR and updating this conditional embedding operator given a new action and observation using kernel Bayes' rule. This is an expressive model: we can model near-arbitrary continuous-valued dynamical systems, limited only by the existence of the conditional embedding operator  $\mathcal{W}_{\mathcal{T}^\circ|\mathcal{T}^\mathcal{A},h_t}$ . In Section 4.2 below we break the PSR state  $\mathcal{W}_{\mathcal{T}^\circ|\mathcal{T}^\mathcal{A},h_t}$  down into its constituent parts, the conditional embedding operators  $\mathcal{C}_{\mathcal{T}^\circ,\mathcal{T}^\mathcal{A}|h_t}$  and  $\mathcal{C}_{\mathcal{T}^\mathcal{A},\mathcal{T}^\mathcal{A}|h_t}$ . We then update these two operators instead of  $\mathcal{W}_{\mathcal{T}^\circ|\mathcal{T}^\mathcal{A},h_t}$  directly. This representation is at least as expressive since we can always reconstruct  $\mathcal{W}_{\mathcal{T}^\circ|\mathcal{T}^\mathcal{A},h_t} = \mathcal{C}_{\mathcal{T}^\circ,\mathcal{T}^\mathcal{A}|h_t} \mathcal{C}_{\mathcal{T}^\mathcal{A},\mathcal{T}^\mathcal{A}|h_t}^{-1}$ .

### 4.1 A Minimal State Space

Instead of working with mean embeddings in generic RKHSs, it is sometimes possible to embed distributions in a finite-dimensional *subspace* of the RKHS. For discrete action-observation PSRs with delta kernels, such a subspace corresponds to a *core set* of tests (Singh et al., 2004). In the more general case, we can *factor* the conditional embedding of the covariance operator  $\mathcal{C}_{\mathcal{T}^\circ,\mathcal{T}^\mathcal{A}|h_t}$  into a finite dimensional operator  $\mathcal{C}_{X^\circ,X^\mathcal{A}|h_t}$  and a conditional covariance operator  $\mathcal{U}$ :

$$\mathcal{C}_{\mathcal{T}^\circ,\mathcal{T}^\mathcal{A}|h_t} = \mathcal{U} \mathcal{C}_{X^\circ,X^\mathcal{A}|h_t} \quad (1)$$

Here  $X = (X^\circ, X^\mathcal{A})$  is a finite set of linear combinations of tests, which we choose to make the factorization possible. Analogous to the discrete case we can find  $\mathcal{U}$  by performing a ‘thin’ (kernel) SVD of the covariance operator  $\mathcal{C}_{(\mathcal{T}^\circ,\mathcal{T}^\mathcal{A})\mathcal{H}}$  and taking the top  $d$  singular vectors as  $\mathcal{U}$ . (So we are choosing  $X$  so that  $\mathcal{U}^* \mathcal{U} = I$ .)

### 4.2 The State Update

To implement the Bayes' rule state update we start with a PSR state at time  $t$ , then take an action, receive an observation, and incorporate this information into the PSR to get the state at time  $t + 1$ . In what follows, we need to be careful about independence between different random variables. For example, if we evaluate  $\phi^{\mathcal{T}^\circ}(\tau_t^\circ)$  and  $\phi^{\mathcal{T}^\circ}(\tau_{t+1}^\circ)$  at the same time step, the realizations will not be independent, even conditioned on the state of the process—if we wanted independent realizations of  $\phi^{\mathcal{T}^\circ}(\tau_t^\circ)$  and  $\phi^{\mathcal{T}^\circ}(\tau_{t+1}^\circ)$ , we'd have to assume the ability to reset the system to a desired state. Below we will take care that we only ask for operators which we can estimate without resets. (If we didn't have to obey this constraint, the algorithm would become somewhat simpler, but the need for resets would constrain applicability.) Therefore, we define several covariance operators which are needed to update the state:  $\mathcal{C}_{\mathcal{T}^\circ,\mathcal{T}^\mathcal{A},\mathcal{H}} \stackrel{\text{def}}{=} \mathbb{E}[\phi^{\mathcal{T}^\circ}(\tau_t^\circ) \otimes \phi^{\mathcal{T}^\mathcal{A}}(\tau_t^\mathcal{A}) \otimes \phi^{\mathcal{H}}(h_t)]$ ,  $\mathcal{C}_{\mathcal{T}^\mathcal{A},\mathcal{T}^\mathcal{A},\mathcal{H}} \stackrel{\text{def}}{=} \mathbb{E}[\phi^{\mathcal{T}^\mathcal{A}}(\tau_t^\mathcal{A}) \otimes \phi^{\mathcal{T}^\mathcal{A}}(\tau_t^\mathcal{A}) \otimes \phi^{\mathcal{H}}(h_t)]$ ,

$$\mathcal{C}_{\mathcal{T}^\circ+, \mathcal{T}^\mathcal{A}+, \mathcal{O}, \mathcal{A}, \mathcal{H}} \stackrel{\text{def}}{=} \mathbb{E}[\phi^{\mathcal{T}^\circ}(\tau_{t+1}^\circ) \otimes \phi^{\mathcal{T}^\mathcal{A}}(\tau_{t+1}^\mathcal{A}) \otimes \phi^\mathcal{O}(o_t) \otimes \phi^\mathcal{A}(a_t) \otimes \phi^{\mathcal{H}}(h_t)], \mathcal{C}_{\mathcal{O}, \mathcal{O}, \mathcal{A}, \mathcal{H}} \stackrel{\text{def}}{=} \mathbb{E}[\phi^\mathcal{O}(o_t) \otimes \phi^\mathcal{O}(o_t) \otimes \phi^\mathcal{A}(a_t) \otimes \phi^{\mathcal{H}}(h_t)], \text{ and } \mathcal{C}_{\mathcal{A}, \mathcal{A}, \mathcal{H}} \stackrel{\text{def}}{=} \mathbb{E}[\phi^\mathcal{A}(a_t) \otimes \phi^\mathcal{A}(a_t) \otimes \phi^{\mathcal{H}}(h_t)].$$

Since  $\mathcal{U}^* \mathcal{C}_{\mathcal{T}^\circ,\mathcal{T}^\mathcal{A}|h_t} = \mathcal{C}_{X^\circ,X^\mathcal{A}|h_t}$  is the characteristic embedding of the probability distribution of *all* tests, we assume that we can compute the embedding of the probability distribution of *any* subset of observations, actions, or tests with a conditional covariance operator from this embedding. For example,

$$\mathcal{C}_{\mathcal{A}, \mathcal{A}|h_t} = \mathcal{W}_{\mathcal{A}, \mathcal{A}|X^\circ, X^\mathcal{A}} \mathcal{C}_{X^\circ, X^\mathcal{A}|h_t}$$

Specifically, we assume that the operators  $\mathcal{W}_{X^{\circ+}, X^{\mathcal{A}+}, \mathcal{O}, \mathcal{A}|X^\circ, X^\mathcal{A}}$ ,  $\mathcal{W}_{\mathcal{T}^\mathcal{A}, \mathcal{T}^\mathcal{A}|X^\circ, X^\mathcal{A}}$ ,  $\mathcal{W}_{\mathcal{O}, \mathcal{O}, \mathcal{A}|X^\circ, X^\mathcal{A}}$ , and  $\mathcal{W}_{\mathcal{A}, \mathcal{A}|X^\circ, X^\mathcal{A}}$  exist and that we can pseudo-invert  $\mathcal{C}_{(X^\circ, X^\mathcal{A})\mathcal{H}}$  such that  $\mathcal{C}_{(X^\circ, X^\mathcal{A})\mathcal{H}} \mathcal{C}_{(X^\circ, X^\mathcal{A})\mathcal{H}}^\dagger = I$ . With some algebraic manipulation it is possible to show that:

$$\begin{aligned} \mathcal{U}^* \mathcal{C}_{(\mathcal{T}^\circ+, \mathcal{T}^\mathcal{A}+, \mathcal{O}, \mathcal{A})\mathcal{H}} (\mathcal{U}^* \mathcal{C}_{(\mathcal{T}^\circ, \mathcal{T}^\mathcal{A})\mathcal{H}})^\dagger &= \mathcal{W}_{X^{\circ+}, X^{\mathcal{A}+}, \mathcal{O}, \mathcal{A}|X^\circ, X^\mathcal{A}} \\ \mathcal{C}_{(\mathcal{T}^\mathcal{A}, \mathcal{T}^\mathcal{A})\mathcal{H}} (\mathcal{U}^* \mathcal{C}_{(\mathcal{T}^\circ, \mathcal{T}^\mathcal{A})\mathcal{H}})^\dagger &= \mathcal{W}_{\mathcal{T}^\mathcal{A}, \mathcal{T}^\mathcal{A}|X^\circ, X^\mathcal{A}} \\ \mathcal{C}_{(\mathcal{O}, \mathcal{O}, \mathcal{A})\mathcal{H}} (\mathcal{U}^* \mathcal{C}_{(\mathcal{T}^\circ, \mathcal{T}^\mathcal{A})\mathcal{H}})^\dagger &= \mathcal{W}_{\mathcal{O}, \mathcal{O}, \mathcal{A}|X^\circ, X^\mathcal{A}} \\ \mathcal{C}_{(\mathcal{A}, \mathcal{A})\mathcal{H}} (\mathcal{U}^* \mathcal{C}_{(\mathcal{T}^\circ, \mathcal{T}^\mathcal{A})\mathcal{H}})^\dagger &= \mathcal{W}_{\mathcal{A}, \mathcal{A}|X^\circ, X^\mathcal{A}} \end{aligned}$$

Using the above equations, we can find covariance operators conditioned on history at time  $t$ , based on the covariance  $\mathcal{C}_{X^{\mathcal{O}}, X^{\mathcal{A}}|h_t}$  at time  $t$ :

$$\begin{aligned}\mathcal{C}_{X^{\mathcal{O}+}, X^{\mathcal{A}+}, \mathcal{O}, \mathcal{A}}|h_t &= \mathcal{W}_{X^{\mathcal{O}+}, X^{\mathcal{A}+}, \mathcal{O}, \mathcal{A}}|X^{\mathcal{O}}, X^{\mathcal{A}}\mathcal{C}_{X^{\mathcal{O}}, X^{\mathcal{A}}|h_t} \\ \mathcal{C}_{\mathcal{T}^{\mathcal{A}}, \mathcal{T}^{\mathcal{A}}|h_t &= \mathcal{W}_{\mathcal{T}^{\mathcal{A}}, \mathcal{T}^{\mathcal{A}}|X^{\mathcal{O}}, X^{\mathcal{A}}\mathcal{C}_{X^{\mathcal{O}}, X^{\mathcal{A}}|h_t} \\ \mathcal{C}_{\mathcal{O}, \mathcal{O}, \mathcal{A}}|h_t &= \mathcal{W}_{\mathcal{O}, \mathcal{O}, \mathcal{A}}|X^{\mathcal{O}}, X^{\mathcal{A}}\mathcal{C}_{X^{\mathcal{O}}, X^{\mathcal{A}}|h_t} \\ \mathcal{C}_{\mathcal{A}, \mathcal{A}}|h_t &= \mathcal{W}_{\mathcal{A}, \mathcal{A}}|X^{\mathcal{O}}, X^{\mathcal{A}}\mathcal{C}_{X^{\mathcal{O}}, X^{\mathcal{A}}|h_t}\end{aligned}$$

Finally, in order to update our state, we execute two instances of KBR. First, when we choose an action  $a_t$ , we update:

$$\begin{aligned}\mathcal{C}_{X^{\mathcal{O}+}, X^{\mathcal{A}+}, \mathcal{O}}|h_t, a_t &= \mathcal{C}_{(X^{\mathcal{O}+}, X^{\mathcal{A}+}, \mathcal{O})\mathcal{A}}|h_t \mathcal{C}_{\mathcal{A}, \mathcal{A}}|h_t^{-1} \phi^{\mathcal{A}}(a_t) \\ \mathcal{C}_{\mathcal{O}, \mathcal{O}}|h_t, a_t &= \mathcal{C}_{(\mathcal{O}, \mathcal{O})\mathcal{A}}|h_t \mathcal{C}_{\mathcal{A}, \mathcal{A}}|h_t^{-1} \phi^{\mathcal{A}}(a_t)\end{aligned}$$

Next, when we receive the observation generated by the system,  $o_t$ , we incorporate it to calculate the joint conditional covariance:

$$\mathcal{C}_{X^{\mathcal{O}+}, X^{\mathcal{A}+}|h_t, a_t, o_t} = \mathcal{C}_{(X^{\mathcal{O}+}, X^{\mathcal{A}+})\mathcal{O}}|h_t, a_t \mathcal{C}_{\mathcal{O}, \mathcal{O}}|h_t, a_t^{-1} \phi^{\mathcal{O}}(o_t)$$

Finally, the joint conditional covariance at time  $t + 1$  is identified as  $\mathcal{C}_{\mathcal{T}^{\mathcal{O}+}, \mathcal{T}^{\mathcal{A}+}|h_t, a_t, o_t}$ :

$$\mathcal{C}_{X^{\mathcal{O}}, X^{\mathcal{A}}|h_{t+1}} \equiv \mathcal{C}_{X^{\mathcal{O}+}, X^{\mathcal{A}+}|h_t, a_t, o_t}$$

The PSR state can now be computed:

$$\begin{aligned}\mathcal{C}_{X^{\mathcal{O}}, X^{\mathcal{A}}|h_{t+1}} &\equiv \mathcal{C}_{X^{\mathcal{O}+}, X^{\mathcal{A}+}|h_t, a_t, o_t} \\ \mathcal{C}_{\mathcal{T}^{\mathcal{A}}, \mathcal{T}^{\mathcal{A}}|h_{t+1}} &= \mathcal{W}_{\mathcal{T}^{\mathcal{A}}, \mathcal{T}^{\mathcal{A}}|X^{\mathcal{O}}, X^{\mathcal{A}}\mathcal{C}_{X^{\mathcal{O}}, X^{\mathcal{A}}|h_{t+1}} \\ \mathcal{W}_{\mathcal{T}^{\mathcal{O}}|\mathcal{T}^{\mathcal{A}}, h_{t+1}} &= \mathcal{U}\mathcal{C}_{X^{\mathcal{O}}, X^{\mathcal{A}}|h_{t+1}} \mathcal{C}_{\mathcal{T}^{\mathcal{A}}, \mathcal{T}^{\mathcal{A}}|h_{t+1}}^{-1}\end{aligned}$$

## 5 Learning

The learning algorithm for Hilbert space embeddings of PSRs is very simple. We estimate the covariance operators  $\mathcal{C}_{\mathcal{T}^{\mathcal{O}}, \mathcal{T}^{\mathcal{A}}, \mathcal{H}}$ ,  $\mathcal{C}_{\mathcal{T}^{\mathcal{A}}, \mathcal{T}^{\mathcal{A}}, \mathcal{H}}$ ,  $\mathcal{C}_{\mathcal{T}^{\mathcal{O}+}, \mathcal{T}^{\mathcal{A}+}, \mathcal{O}, \mathcal{A}, \mathcal{H}}$ ,  $\mathcal{C}_{\mathcal{O}, \mathcal{O}, \mathcal{A}, \mathcal{H}}$ , and  $\mathcal{C}_{\mathcal{A}, \mathcal{A}, \mathcal{H}}$  from our observable data, and plug into the above equations to learn the PSR model and to filter in the learned model. If the RKHS is *infinite*, which is often the case, then it is not possible to store or manipulate the above covariances directly. Instead, we use the “kernel trick” and represent all of the covariances and compute state updates with Gram matrices. Further details, including the Gram matrix formulation of the algorithm, can be found in (Boots, 2012).

We expect to have experimental results comparing to previous approaches like Hilbert Space Embeddings of HMMs and POMDPs by the time that the workshop is held.

## References

- Baker, C. (1973). Joint measures and cross-covariance operators. *Transactions of the American Mathematical Society*, 186, 273–289.
- Boots, B. (2012). *Spectral Approaches to Learning Predictive Representations* (PhD. Thesis). Carnegie Mellon University.
- Boots, B., Siddiqi, S. M., & Gordon, G. J. (2010). Closing the learning-planning loop with predictive state representations. *Proceedings of Robotics: Science and Systems VI*.
- Fukumizu, K., Song, L., & Gretton, A. (2011). Kernel bayes’ rule. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira and K. Weinberger (Eds.), *Advances in neural information processing systems 24*, 1737–1745.
- Grunewalder, S., Lever, G., Baldassarre, L., Pontil, M., & Gretton, A. (2012). Modelling transition dynamics in mdps with rkhs embeddings. *CoRR, abs/1206.4655*.
- Littman, M., Sutton, R., & Singh, S. (2002). Predictive representations of state. *Advances in Neural Information Processing Systems (NIPS)*.
- Nishiyama, Y., Boularias, A., Gretton, A., & Fukumizu, K. (2012). Hilbert space embeddings of pomdps. .
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77, 257–285.
- Singh, S., James, M., & Rudary, M. (2004). Predictive state representations: A new theory for modeling dynamical systems. *Proc. UAI*.
- Smola, A., Gretton, A., Song, L., & Schölkopf, B. (2007). A Hilbert space embedding for distributions. *Algorithmic Learning Theory*. Springer.
- Sondik, E. J. (1971). *The optimal control of partially observable Markov processes* PhD. Thesis). Stanford University.
- Song, L., Boots, B., Siddiqi, S. M., Gordon, G. J., & Smola, A. J. (2010). Hilbert space embeddings of hidden Markov models. *Proc. 27th Intl. Conf. on Machine Learning (ICML)*.
- Song, L., Huang, J., Smola, A., & Fukumizu, K. (2009). Hilbert space embeddings of conditional distributions. *International Conference on Machine Learning*.
- Sriperumbudur, B., Gretton, A., Fukumizu, K., Lanckriet, G., & Schölkopf, B. (2008). Injective Hilbert space embeddings of probability measures. .